The Annals of Applied Statistics
2018, Vol. 12, No. 3, 1558–1582
https://doi.org/10.1214/17-AOAS1121

ADAPTIVE-WEIGHT BURDEN TEST FOR ASSOCIATIONS BETWEEN QUANTITATIVE TRAITS AND GENOTYPE DATA WITH COMPLEX CORRELATIONS

BY XIAOWEI WU*, TING GUAN*, DAJIANG J. LIU†, LUIS G. LEÓN NOVELO‡ AND DIPANKAR BANDYOPADHYAY§,[1]

Virginia Tech, Pennsylvania State University College of Medicine†, University of Texas Health Science Center‡ and Virginia Commonwealth University§*

High throughput sequencing has often been used to screen samples from pedigrees or with population structure, producing genotype data with complex correlations caused by both familial relation and linkage disequilibrium. With such data it is critical to account for these genotypic correlations when assessing the contribution of multiple variants by gene or pathway. Recognizing the limitations of existing association testing methods, we propose *Adaptive-weight Burden Test* (ABT), a retrospective, mixed model test for genetic association of quantitative traits on genotype data with complex correlations. This method makes full use of genotypic correlations across both samples and variants and adopts "data driven" weights to improve power. We derive the ABT statistic and its explicit distribution under the null hypothesis and demonstrate through simulation studies that it is generally more powerful than the fixed-weight burden test and family-based SKAT in various scenarios, controlling for the type I error rate. Further investigation reveals the connection of ABT with kernel tests, as well as the adaptability of its weights to the direction of genetic effects. The application of ABT is illustrated by a gene-based association analysis of fasting glucose using data from the NHLBI "Grand Opportunity" Exome Sequencing Project.

1. Introduction. Next generation sequencing technologies have been undergoing rapid evolution in recent years, enabling high resolution genotyping in a fast, efficient, and cost effective way [Ansorge (2009), Shendure and Ji (2008)]. These technologies, when applied to family or population based samples, produce rich resources of genotype data with complex correlations, that is, they are correlated across both samples and variants due to familial relation (or population stratification) and linkage disequilibrium (LD), respectively. Though such genotypic correlations could potentially benefit many applications—including data imputation, quality control, and functional annotation—how to effectively use this information for assessing genetic associations with complex diseases remains a challenge.

Received May 2016; revised August 2017.

[1]Supported in part by Grants R03DE023372 and R01DE024984 from the NIH/NIDCR, and VCU Massey Cancer Center Support Grant P30CA016059 from NIH/NCI.

Key words and phrases. Genetic association test, burden test, kernel test, adaptive weight, complex genotypic correlation.

In genome-wide association studies (GWASs), two types of tests are commonly used depending on whether single-variant effect or joint effect of multiple variants in a predefined genomic region is of interest. Compared to single-variant association tests, multiple-variant association tests (i.e., SNP-set association tests) are believed to be advantageous in that they are more powerful in aggregating small signals from each single variant, especially when the minor allele frequency is low [Asimit and Zeggini (2010)], capturing possible interactions among variants [Ma, Clark and Keinan (2013)], reducing multiple testing burden [Wu et al. (2010)], and leading to interpretable results targeted to genes or LD haplotype blocks [Li et al. (2011)].

A number of multiple-variant association tests have been proposed. They generally fall under two broad categories: "burden tests" and "kernel tests." Burden tests group variants into a single variable called the genetic burden score by some transformations or projections and then perform association testing on the burden score. Typical collapsing methods, including rare variant indicator or weighted sum, have been developed for both unrelated [Li and Leal (2008), Madsen and Browning (2009), Morgenthaler and Thilly (2007), Price et al. (2010a)] and related [Chen, Meigs and Dupuis (2013), Schaid et al. (2013)] individuals. Other dimension reduction techniques, such as Fourier transformation [Wang and Elston (2007)], principal component analysis [Gauderman et al. (2007), Wang and Abbott (2008)], and partial least-squares regression [Chun et al. (2011)], have also been applied in grouping multiple variants. Although easy to implement, burden tests rely largely on the assumption of homogeneity in the magnitude and direction of the genetic effects from individual variants. When the genomic region of interest includes both risk (positively associated) and protective (negatively associated) variants, or, when inappropriate weights (contradicting the true genetic effects) are used, burden tests may experience loss of power.

Recently, kernel tests have been receiving increased attention in GWASs. With its roots in kernel machine regression [Wu et al. (2010)] and mixed effects model, kernel tests adopt a statistic of quadratic form, which essentially is a weighted sum of the score statistics for testing individual variant effects [Wu et al. (2011)]. Methods in this broad category include the C-alpha test [Neale et al. (2011)], SKAT [Wu et al. (2011)], and LSKM [Kwee et al. (2008)] for unrelated individuals and have been extended to allow related individuals [Chen, Meigs and Dupuis (2013), Schaid et al. (2013), Schifano et al. (2012), Wang et al. (2013a, 2013b)]. In particular, this approach assumes random effects for individual variants and tests the regression coefficients of the variants by a variance component score test. Since the aggregation is on individual score statistics instead of on variants, kernel tests allow different directions and magnitudes of effects for individual variants. It has been shown that burden tests are more powerful when the variants to be tested are most causal with effects of the same direction and similar magnitudes, whereas kernel tests are more powerful when the effects of causal variants are in differ-

ent directions or a large proportion of neutral variants present [Chen, Meigs and Dupuis (2013), Lee, Wu and Lin (2012), Schaid et al. (2013), Wang, Chen and Yang (2012), Wu et al. (2011)]. To borrow strength from both approaches and avoid loss of power in certain scenarios, it is possible to combine test statistics from the two categories [Jiang and McPeek (2013), Lee, Wu and Lin (2012), Lee et al. (2012)], albeit determining the optimal combination weight may be problematic in real data applications, and the null distribution of the combined statistic is usually hard to derive.

In the category of burden tests, considerable effort have been made to seek weighting strategies that allow for the presence of both risk and protective variants [Fang, Zhang and Sha (2014), Han and Pan (2010), Lin and Tang (2011), Liu and Leal (2010), Sha and Zhang (2014), Sha et al. (2012)]. However, as the optimal weights are functions of the genotype data, the resulting test statistic does not follow χ_1^2 null distribution as in fixed-weight score-based burden tests. None of these methods derive the explicit null distribution after employing the optimal weights, but use permutations to evaluate the p-value instead. Since permutations are computationally expensive for whole genome analysis and also not straightforward when the samples include related individuals, the application of these permutation-based methods is largely restricted in GWASs with pedigrees. Moreover, because all these burden tests use prospective regression (which considers trait as random response and genotype data as fixed predictors), the LD correlation among the considered variants is often hard to be incorporated into the weights.

In this paper, we focus on association mapping of quantitative traits on genotype data with complex correlations. Specifically, we try to answer the following questions:

(1) How to model the genotypic correlations across both samples and variants simultaneously in an efficient way?

(2) How to choose the optimal weights of a burden test in the presence of complex genotypic correlations?

(3) Is there a connection between burden tests and kernel tests, under certain weighting strategies?

(4) How are the optimal weights adaptive to the direction of individual variant effects?

To address these, we propose the *Adaptive-weight Burden Test* (ABT), a retrospective, mixed-model test, which incorporates complex genotypic correlations and adopts data driven weights to improve power. We show that the explicit null distribution of the ABT can be obtained by appropriately projecting genotype data and combining independent individual score tests. Therefore, compared to other permutation-based optimal tests, ABT is computationally more efficient.

The rest of the paper is organized as follows. Section 2 presents the relevant background of genetic association testing and some preliminaries regarding the single variant MASTOR test [Jakobsdottir and McPeek (2013)], including its ex-

tension to a retrospective, fixed-weight burden test. Based on these preliminaries, we introduce the statistical framework of the ABT test, sketch some of its theoretical properties and illuminate its connection to the kernel tests. In Section 3, we demonstrate via synthetic data that the proposed ABT test is generally more powerful than the fixed-weight burden test and family-based SKAT under various scenarios, well-controlling for the type I error. Furthermore, the weights of ABT are able to adapt to the direction of individual variant effects. In Section 4, we illustrate the use of our ABT test by a gene-based association analysis of fasting glucose using data from the NHLBI "Grand Opportunity" Exome Sequencing Project. Finally, in Section 5, we present some concluding remarks.

2. Statistical framework of the ABT. Consider an association study where a group of n individuals are sampled for phenotype, covariate and genotype data. For simplicity, we do not assume any missingness in the data (i.e., each individual is assumed to have complete data in phenotypes, covariates and genotypes), though all the results hereafter can be extended to the incomplete data case in a way similar to Jakobsdottir and McPeek (2013). The phenotype data are collected for a quantitative trait and denoted by a vector $\boldsymbol{Y}$ of length n. The covariates form an $n \times k$ matrix $\boldsymbol{Z}$, with the columns representing k nongenetic variables (intercept included) such as age, sex and body mass index (BMI). We consider testing for association between the quantitative trait and a genetic region of m variant sites. Each typed variant is assumed to be biallelic, with the alleles arbitrarily labeled as "0" and "1." So, the genotype data can be written as an $n \times m$ matrix $\boldsymbol{G} = [\boldsymbol{G}_1, \boldsymbol{G}_2, \ldots, \boldsymbol{G}_m]$ with the (i, j)th element coded as $G_{ij} = \frac{1}{2} \times$ (the number of alleles of type 1 in individual i at variant site j), $1 \le i \le n$, $1 \le j \le m$. These m variants are further assumed to have a certain LD structure, with the correlation matrix defined as

$$\boldsymbol{R} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{12} & 1 & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1m} & r_{2m} & \cdots & 1 \end{pmatrix},$$

where $r_{ij} = (p_{11} - p_i p_j)/\sqrt{p_i(1-p_i)p_j(1-p_j)}$, $1 \le i \ne j \le m$, is the correlation coefficient between variant i and variant j, p_i and p_j are the allele frequencies of variants i, j respectively, and p_{11} is the frequency of the haplotype having alleles 1 at both variants. In addition to the correlation among genetic markers, we also consider the correlation (i.e., relatedness) among sampled individuals in this current work. We assume a known pedigree structure for the sampled individuals and define the kinship matrix as

$$\boldsymbol{\Phi} = \begin{pmatrix} 1+h_1 & 2\phi_{12} & \cdots & 2\phi_{1n} \\ 2\phi_{12} & 1+h_2 & \cdots & 2\phi_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 2\phi_{1n} & 2\phi_{2n} & \cdots & 1+h_n \end{pmatrix},$$

where h_i is the inbreeding coefficient of individual i, and ϕ_{ij} is the kinship coefficient between individuals i and $j, 1 \le i, j \le n$. For outbred individuals, the kinship matrix can be considered as the correlation matrix among individual genotypes. As a special case, for unrelated individuals, the kinship matrix reduces to the identity matrix.

To conveniently model genotypic correlations caused by both familial relation and LD, we will treat genotypes as random and conduct a retrospective analysis based on $\boldsymbol{G}|(\boldsymbol{Y}, \boldsymbol{Z})$. This approach, originated from the MQLS [Thornton and McPeek (2007)] and MASTOR [Jakobsdottir and McPeek (2013)] tests, is different from most existing association testing methods for related individuals, such as the MONSTER [Jiang and McPeek (2013)], family-based burden test, and family-based SKAT [Chen, Meigs and Dupuis (2013)], which are all based on a prospective model $\boldsymbol{Y}|(\boldsymbol{G}, \boldsymbol{Z})$. The next few subsections proceed as follows. For a better understanding, we begin with a brief introduction to the single variant MASTOR test and extend it to a retrospective, fixed-weight burden test. Next, we derive the ABT statistic and illuminate its connection with kernel tests.

2.1. *MASTOR for single variant effect.* In single variant analysis, MASTOR [Jakobsdottir and McPeek (2013)] is a retrospective, quasi-likelihood score test for genetic association of a quantitative trait in samples with related individuals. MASTOR is able to gain additional power by making full use of the sample relationship information to incorporate partially missing data, therefore, it is more advantageous than other competitors. Considering a biallelic genetic variant $\boldsymbol{X}$ of interest (an example in our general setting described above is to let $\boldsymbol{X} = \boldsymbol{G}_j, 1 \le j \le m$), the MASTOR test statistic (for complete data) takes the form

$$S_{\mathrm{MAS}} = \frac{(\boldsymbol{V}^T \boldsymbol{X})^2}{\widehat{\mathrm{Var}_0}(\boldsymbol{V}^T \boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{Z})}, \tag{2.1}$$

where $\boldsymbol{V} = \widehat{\boldsymbol{\Sigma}}_0^{-1}(\boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{\beta}}_0)$ is the transformed phenotypic residual obtained from the model $\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_0)$. Here, $\boldsymbol{\beta}_0$ represents the regression coefficients under the null hypothesis of no genetic association, and $\boldsymbol{\Sigma}_0$ is the trait covariance matrix under the null, usually taking a variance component form $\sigma_e^2 \boldsymbol{I} + \sigma_a^2 \boldsymbol{\Phi}$. Let $\boldsymbol{P} = \widehat{\boldsymbol{\Sigma}}_0^{-1} - \widehat{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{Z}(\boldsymbol{Z}^T \widehat{\boldsymbol{\Sigma}}_0^{-1} \boldsymbol{Z})^{-1} \boldsymbol{Z}^T \widehat{\boldsymbol{\Sigma}}_0^{-1}$. $\boldsymbol{V}$ is often short notated as $\boldsymbol{PY}$. Under the assumption of the retrospective model that $\mathrm{Var}_0(\boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{Z}) = \sigma_X^2 \boldsymbol{\Phi}$, where σ_X^2 is an unknown scalar representing the variance of $\boldsymbol{X}$, Equation (2.1) can be written as

$$S_{\mathrm{MAS}} = \frac{(\boldsymbol{V}^T \boldsymbol{X})^2}{(\boldsymbol{V}^T \boldsymbol{\Phi} \boldsymbol{V})\widehat{\sigma}_X^2} = \frac{(\boldsymbol{Y}^T \boldsymbol{P} \boldsymbol{X})^2}{(\boldsymbol{Y}^T \boldsymbol{P} \boldsymbol{\Phi} \boldsymbol{P} \boldsymbol{Y})\widehat{\sigma}_X^2}. \tag{2.2}$$

When the Hardy–Weinberg equilibrium (HWE) is assumed at the variant, a simple estimator of σ_X^2 can be obtained as $\widehat{\sigma}_X^2 = \widehat{p}(1 - \widehat{p})/2$, where $\widehat{p} = (\boldsymbol{1}^T \boldsymbol{\Phi}^{-1} \boldsymbol{1})^{-1} \boldsymbol{1}^T \boldsymbol{\Phi}^{-1} \boldsymbol{X}$ is the best linear unbiased estimator (BLUE) [McPeek, Wu and Ober (2004)] of the allele frequency p of $\boldsymbol{X}$, and $\boldsymbol{1}$ denotes a vector

with every element equal to 1. In practice, a more general and robust estimator $\widehat{\sigma}_X^2 = \boldsymbol{X}^T \boldsymbol{U} \boldsymbol{X}/(n-1)$ can be used instead [Thornton and McPeek (2010)], where $\boldsymbol{U} = \boldsymbol{\Phi}^{-1} - \boldsymbol{\Phi}^{-1}\mathbf{1}(\mathbf{1}^T\boldsymbol{\Phi}^{-1}\mathbf{1})^{-1}\mathbf{1}^T\boldsymbol{\Phi}^{-1}$. Under the null, the MASTOR statistic follows a χ_1^2 distribution.

2.2. *Retrospective, fixed-weight burden test.* MASTOR can be extended to multiple-variant testing, that is, to test association between trait and a set of genetic variants. An easy extension is through burden tests, which are constructed following a two-step procedure—first, collapse multiple variants into a genetic burden score by linear combination, and then obtain the test statistic similarly as in single-variant tests. Following this formulation, we introduce a *fixed-weight burden test* for association between a quantitative trait and genotype data with LD and sample relatedness. We acronymize this method by FBT, where the F stands for "fixed-weight" or "family-based," yet, in the latter sense, this retrospective burden test is different from the prospective *famBT* method of Chen, Meigs and Dupuis (2013). Fixed-weight here refers to the setting of prescribed weights, in contrast to the weights derived in the next subsection which are data adaptive. Contrary to other burden tests, FBT is based on a retrospective model analogous to MASTOR and also possesses nice properties, such as ability to incorporate partially missing data and robustness to misspecification of the phenotype model. A similar method can be found in Schaid et al. (2013), with a different approach adopted for defining residuals in the null trait model and deriving covariances of the genetic burden score.

For genotype data $\boldsymbol{G}$ consisting of m variants with corresponding allele frequencies $p_1, p_2, \ldots, p_m$, we consider the weighted-sum burden score

$$\boldsymbol{X} = \sum_{i=1}^{m} w_i \boldsymbol{G}_i = \boldsymbol{G}\boldsymbol{W}, \tag{2.3}$$

where $\boldsymbol{W} = [w_1, w_2, \ldots, w_m]^T$ is a prescribed weight vector. Following the MASTOR statistic (2.1), a fixed-weight burden test statistic based on $\boldsymbol{X}$ can be constructed as

$$S_{\mathrm{FBT}} = \frac{(\boldsymbol{V}^T\boldsymbol{X})^2}{\boldsymbol{V}^T\widehat{\boldsymbol{\Sigma}}_X\boldsymbol{V}}, \tag{2.4}$$

where $\widehat{\boldsymbol{\Sigma}}_X$ is an estimator of the covariance matrix of $\boldsymbol{X}$ under the null. Analogous to MASTOR, it can be shown that conditioning on $\boldsymbol{W}, \boldsymbol{Y}, \boldsymbol{Z}$, S_{FBT} follows a χ_1^2 distribution under the null hypothesis that the genetic score $\boldsymbol{X}$ is not associated with $\boldsymbol{Y}$, that is, the set of variants $\boldsymbol{G}$ has no genetic effects on $\boldsymbol{Y}$ after collapsed into $\boldsymbol{X}$. Furthermore, if the vectorized $\boldsymbol{G}$ has a Kronecker product covariance structure, then $\boldsymbol{\Sigma}_X$ can be expressed in terms of the weight vector, across-column covariance, and across-row correlation as $\boldsymbol{\Sigma}_X = (\boldsymbol{W}^T\boldsymbol{D}\boldsymbol{R}\boldsymbol{D}\boldsymbol{W})\boldsymbol{\Phi}$, where

$\boldsymbol{D} = \text{diag}\{\sigma_j\}$, $1 \le j \le m$ and σ_j is the marginal standard deviation of variant $\boldsymbol{G}_j$ (see Section 1 of the supplemental article [Wu et al. (2018)]).

Correspondingly, if $\boldsymbol{R}$ and $\boldsymbol{\Phi}$ are assumed known, an appropriate estimator of $\boldsymbol{\Sigma}_X$ is

$$\widehat{\boldsymbol{\Sigma}}_X = (\boldsymbol{W}^T \widehat{\boldsymbol{D}} \boldsymbol{R} \widehat{\boldsymbol{D}} \boldsymbol{W}) \boldsymbol{\Phi}, \tag{2.5}$$

where the jth diagonal element of $\widehat{\boldsymbol{D}}$ is estimated as $\widehat{\sigma}_j = \sqrt{\widehat{p}_j(1-\widehat{p}_j)/2}$, $1 \le j \le m$ under HWE, with $\widehat{p}_j$, the BLUE of the allele frequency p_j as previously defined. Therefore, the FBT statistic becomes

$$S_{\text{FBT}} = \frac{\boldsymbol{W}^T \boldsymbol{G}^T \boldsymbol{V} \boldsymbol{V}^T \boldsymbol{G} \boldsymbol{W}}{[\boldsymbol{W}^T (\widehat{\boldsymbol{D}} \boldsymbol{R} \widehat{\boldsymbol{D}}) \boldsymbol{W}](\boldsymbol{V}^T \boldsymbol{\Phi} \boldsymbol{V})}. \tag{2.6}$$

Clearly, S_{FBT} is invariant to the scale of $\boldsymbol{W}$. As a special case, when $m = 1$, S_{FBT} in (2.6) reduces to S_{MAS} in (2.2).

In general, the complex correlation structure in the genotype data, that is, both $\boldsymbol{R}$ and $\boldsymbol{\Phi}$ in (2.6), is not assumed known a priori and needs to be estimated from the data. For simplicity, in our analysis we assume that the across-row correlation $\boldsymbol{\Phi}$ (or the entire pedigree structure) is known. We note that assuming $\boldsymbol{\Phi}$ to be known is reasonable since the information of sample relatedness is often available for most family-based studies [Splansky et al. (2007)]. When population structure or cryptic relatedness presents, $\boldsymbol{\Phi}$ can be estimated from genome screen data [Thornton and McPeek (2010)]. As for the across-column correlation $\boldsymbol{R}$, one may obtain its estimate $\widehat{\boldsymbol{R}}$ from a reference population, for example, the one provided by the 1000 Genomes Project [The 1000 Genomes Project Consortium (2010)]. When the reference panel information is not available, a simple and practical way to calculate $\widehat{\boldsymbol{R}}$ from $\boldsymbol{G}$ is to use the sample correlation of the matrix $\boldsymbol{\Phi}^{-1/2}\boldsymbol{G}$. This estimation, however, has a nonnegligible impact on the performance of the test statistic in terms of type I error and power. This issue will be further discussed in Section 3.3.

2.3. *Adaptive-weight burden test.* A common problem in existing burden tests is to have the variant weights prespecified or set in some ad hoc way, thereby lacking theoretical justification for the test to be statistically powerful. For example, one may use a simple sum method to assign equal weights to the variants. A better strategy is the weighted sum method [Madsen and Browning (2009)], which assigns weights according to allele frequency, that is, $w_j \propto 1/\sqrt{\widehat{p}_j(1-\widehat{p}_j)}$. It has been shown that with prescribed weights, burden tests cannot distinguish the effects from risk and protective variants, that is, when the minor alleles across all sites have effects in different directions—some positive and some negative [Chen, Meigs and Dupuis (2013), Schaid et al. (2013), Wu et al. (2011)]. To overcome this deficiency, several adaptive burden tests have been proposed [Fang, Zhang and Sha (2014), Han and Pan (2010), Lin and Tang (2011), Liu and Leal (2010), Sha and Zhang (2014), Sha et al. (2012)], however, most of these weighting strategies are

empirical and cannot make full use of the complex correlation information. Hence, although easy to implement, burden tests are usually not preferred in real applications where no a priori information exists on the effect of individual variants. This motivates us to look for a burden test that is able to "let the data speak for themselves," that is, with weights adaptive to the direction of individual variant effects.

Let $\boldsymbol{A} = \widehat{\boldsymbol{D}}\boldsymbol{R}\widehat{\boldsymbol{D}}$ and $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}^T$ where $\boldsymbol{b} = \boldsymbol{G}^T\boldsymbol{V}$. From the generalized Rayleigh quotient form of (2.6), we can show that the weight vector $\boldsymbol{W}^*$ that maximizes S_{FBT} satisfies

$$\boldsymbol{W}^* \propto \boldsymbol{A}^{-1}\boldsymbol{b} = (\widehat{\boldsymbol{D}}\boldsymbol{R}\widehat{\boldsymbol{D}})^{-1}\boldsymbol{G}^T\boldsymbol{V}. \tag{2.7}$$

Another representation of $\boldsymbol{W}^*$ is a vector proportional to (or, with the same direction as) the eigenvector of $\boldsymbol{A}^{-1}\boldsymbol{B}$ corresponding to the largest eigenvalue. We refer to burden tests with such weights as ABT and denote the maximized burden statistic by S_{ABT}. It follows by plugging in $\boldsymbol{W}^*$ to (2.6) that

$$S_{\mathrm{ABT}} = \frac{\boldsymbol{V}^T\boldsymbol{G}(\widehat{\boldsymbol{D}}\boldsymbol{R}\widehat{\boldsymbol{D}})^{-1}\boldsymbol{G}^T\boldsymbol{V}}{\boldsymbol{V}^T\boldsymbol{\Phi}\boldsymbol{V}}. \tag{2.8}$$

The key question now pertains to derivation of the null distribution of S_{ABT}.

Note, maximizing S_{FBT} enables the weights to accommodate to the direction of individual variant effects, however, this optimization may not necessarily lead to maximal power. Since $\boldsymbol{W}^*$ itself is a function of $\boldsymbol{G}$, S_{ABT} may not follow a χ^2_1 distribution under the null hypothesis. One may attempt to obtain its null distribution through integrating out $\boldsymbol{W}^*$ by $\int f_0(S_{\mathrm{FBT}}|\boldsymbol{W} = \boldsymbol{W}^*)\,dF(\boldsymbol{W}^*)$, where f_0 is the PDF of χ^2_1—the null density of S_{FBT}, and F denotes the distribution function of $\boldsymbol{W}^*$ determined by (2.7). We notice that the term $\boldsymbol{G}(\widehat{\boldsymbol{D}}\boldsymbol{R}\widehat{\boldsymbol{D}})^{-1/2}$ in the numerator of (2.8) is a transformed genotype matrix of m genetic variants with only across-row correlations. Therefore, by matrix algebra, S_{ABT} can be considered as the summation of the MASTOR statistics from m independent variants (after appropriate projection to eliminate LD correlations and standardize marginal variances), and hence follows χ^2_m distribution under the null hypothesis. This finding greatly simplifies the p-value calculation as compared to other permutation-based approaches [Fang, Zhang and Sha (2014), Han and Pan (2010), Lin and Tang (2011), Liu and Leal (2010), Sha and Zhang (2014), Sha et al. (2012)] and makes ABT suitable to real applications of genome-wide analysis. When $\boldsymbol{R}$ is unknown in (2.8), we replace it with an estimate from $\boldsymbol{G}$. We will show through a connection with kernel tests that in such a case the null distribution of S_{ABT} can be determined by a mixture of χ^2_1's.

As a simplified example using the configuration of unrelated individuals and common variants (see Scenario I in Section 3.1), Figure 1 demonstrates the empirical cumulative distribution functions (ECDFs) of S_{ABT} and S_{FBT} obtained from 10,000 simulated replicates under the null. The simulation is based on $n = 1600$ individuals and a varying number of variants ($m = 10, 50, 100$), with moderate

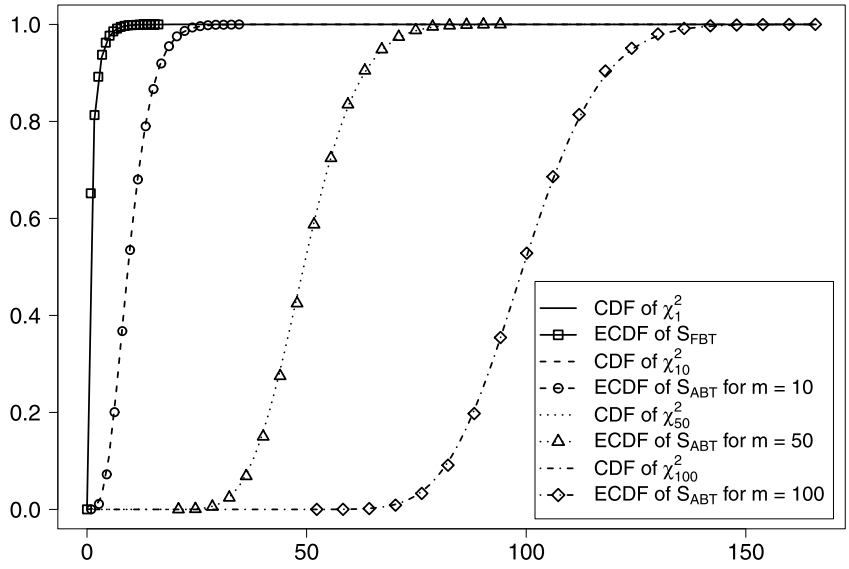

FIG. 1. *Empirical CDFs* (*shown in different markers*) *of* S_{ABT} *and* S_{FBT} *based on* 10,000 *simulated replicates under the null hypothesis and their theoretical CDFs* (*shown in different line types*).

correlation (0.5) between the latent standard normal random variables which are used to generate haplotypes in LD by dichotomization. The details are available in Section 3.1.

2.4. *ABT as a kernel test with generalized Madsen–Browning weights*. The novel finding of the null distribution of S_{ABT} illustrated in the previous subsection provides us the motivation to further explore the relation between ABT and family-based kernel tests. Kernel tests (also called "quadratic tests" or "variance component tests") assume random effects for the regression coefficients of multiple variants. Specifically, family-based kernel tests under our study design is based on a prospective model

$$\boldsymbol{Y} = \boldsymbol{Z\beta} + \boldsymbol{G\gamma} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N\big(\boldsymbol{0}, \sigma_e^2 \boldsymbol{I} + \sigma_a^2 \boldsymbol{\Phi}\big). \tag{2.9}$$

In this model, $\boldsymbol{\gamma}$ is a vector of the genetic effects, and its jth component follows a distribution with mean 0 and variance $w_j^2\tau$, $1 \le j \le m$. Here, w_j is a fixed, prescribed weight for the jth variant effect. Testing $\boldsymbol{\gamma} = \boldsymbol{0}$ is equivalent to testing the common variance component $\tau = 0$. Following this formulation, Chen, Meigs and Dupuis (2013) developed the famSKAT statistic for testing association of a quantitative trait in family samples as

$$S_{\mathrm{KT}} = \boldsymbol{V}^T \boldsymbol{G}\boldsymbol{W}\boldsymbol{W}\boldsymbol{G}^T \boldsymbol{V}, \tag{2.10}$$

where $\boldsymbol{W} = \mathrm{diag}\{w_j\}$, $1 \le j \le m$.

Under the null hypothesis, S_{KT} follows the distribution of $\sum_{i=1}^{m} \lambda_i \chi_{1,i}^2$ where λ_i's are the eigenvalues of the matrix $\boldsymbol{WG}^T \boldsymbol{PGW}$, and $\chi_{1,i}^2$'s are independent χ_1^2 random variables. We note that the $\boldsymbol{W}$ matrix in the above formula plays the role of the square root of the $\boldsymbol{W}$ matrix defined in Chen, Meigs and Dupuis (2013), and the $\boldsymbol{P}$ matrix in our notation is connected to the $\boldsymbol{P}_0$ matrix defined in Chen, Meigs and Dupuis (2013) by $\boldsymbol{P} = \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{P}_0 \widehat{\boldsymbol{\Sigma}}^{-1}$. Comparing the ABT statistic (2.8) with the famSKAT statistic (2.10), we see that ABT, though straightforwardly derived from the fixed-weight burden test under retrospective setting, can be treated formally as a kernel test with the weight matrix

$$\boldsymbol{W}^{\#} = (\boldsymbol{V}^T \boldsymbol{\Phi} \boldsymbol{V})^{-1/2} (\widehat{\boldsymbol{D}} \boldsymbol{R} \widehat{\boldsymbol{D}})^{-1/2}. \tag{2.11}$$

This interesting finding shows that, using data-driven weights selected under the guidance of complex correlations $\boldsymbol{\Phi}$ and $\boldsymbol{R}$ in the genotype data, burden tests and kernel tests reach a formal equivalence, regardless of fixed or random effect models, and the underlying prospective or retrospective models on which they are based.

From the kernel test perspective, the weight matrix $\boldsymbol{W}^{\#} \propto (\widehat{\boldsymbol{D}} \boldsymbol{R} \widehat{\boldsymbol{D}})^{-1/2}$. The diagonal components of this matrix, when all variants in the genetic region to be tested are in linkage equilibrium, that is, $\boldsymbol{R} = \boldsymbol{I}$, become the commonly used Madsen–Browning weights [Madsen and Browning (2009)], that is, $w_j \propto 1/\sqrt{\widehat{p}_j(1-\widehat{p}_j)}$. We therefore call $\boldsymbol{W}^{\#}$ in equation (2.11) the generalized Madsen–Browning (GMB) weights. We note that the GMB weights refer to the entire matrix of $\boldsymbol{W}^{\#}$, not just its diagonal components, because the weight of an individual variant statistic should also be affected by the weights of other variant statistics on linked sites in the presence of LD. Another analogous view of the ABT statistic to the famSKAT statistic is through a two-step calculation. First, eliminate LD correlations in the genotype matrix $\boldsymbol{G}$ by right multiplying $(\widehat{\boldsymbol{D}} \boldsymbol{R} \widehat{\boldsymbol{D}})^{-1/2}$, and then obtain a weighted sum of the individual score statistics for the transformed variants of the decorrelated genotype matrix $\boldsymbol{G}(\widehat{\boldsymbol{D}} \boldsymbol{R} \widehat{\boldsymbol{D}})^{-1/2}$. The weight matrix used in the second step, denoted as $\widetilde{\boldsymbol{W}}$, is a diagonal matrix given by $\widetilde{\boldsymbol{W}} = (\boldsymbol{V}^T \boldsymbol{\Phi} \boldsymbol{V})^{-1/2} \boldsymbol{I}$. This indicates that ABT is indeed a kernel test on the decorrelated genotype matrix, using identical weights equal to $(\boldsymbol{V}^T \boldsymbol{\Phi} \boldsymbol{V})^{-1/2}$. It is not surprising to see that the weights $\widetilde{\boldsymbol{W}}$ of these individual score statistics do not depend on the marginal variances because the decorrelated genotype matrix $\boldsymbol{G}(\widehat{\boldsymbol{D}} \boldsymbol{R} \widehat{\boldsymbol{D}})^{-1/2}$ has an identity covariance matrix.

Similar to kernel tests, we may obtain an alternative for the null distribution of ABT as $\sum_{i=1}^{m} \lambda_i \chi_{1,i}^2$, where λ_i's are the eigenvalues of the matrix $\boldsymbol{W}^{\#} \boldsymbol{G}^T \boldsymbol{PG} \boldsymbol{W}^{\#}$ and $\chi_{1,i}^2$'s are independent χ_1^2 random variables. We will call χ_m^2 the theoretical null distribution of ABT and call the mixture of $\chi_{1,i}^2$'s the practical null distribution of ABT. The latter is usually used when the complex genotypic covariance structure does not follow a Kronecker product form, that is, $\boldsymbol{\Phi}$ and $\boldsymbol{R}$ are not separable [Fuentes (2006)], and $\boldsymbol{R}$ is unknown and needs to be estimated from $\boldsymbol{G}$.

These two null distributions and their impact on the performance of ABT will be further explored in Section 3.3. Since ABT is derived from the burden test framework, yet takes a form of kernel tests, we expect it to be more appropriate than the unified method of linear-combining fixed-weight burden and kernel test statistics as $aS_{\mathrm{FBT}} + (1-a)S_{\mathrm{KT}}$, $0 \le a \le 1$ [Jiang and McPeek (2013), Lee, Wu and Lin (2012)]. Further evidence can be found in our simulation results in Section 3.4.

3. Simulation studies. In this section, we perform simulations to assess the type I error rate of ABT and compare its power to that of FBT, family-based kernel test (abbreviated as KT) and MONSTER [Jiang and McPeek (2013)]. As a by-product of these simulations, we also illustrate that the weights of ABT can adapt to the direction of the true genetic effects.

3.1. *Data generation scenarios*. We simulate data under nine genotypic scenarios (I–IX) using different variant (common/rare/mixed) and individual (unrelated/related/mixed) settings, as shown in Table 1.

In these simulations, we generate genotype data for 1600 individuals and a varying number of variants with minor allele frequencies (MAFs) sampled independently from uniform distribution. For common (Scenarios I, IV, VII) and rare variants (Scenarios II, V, VIII), the MAFs are sampled from $\mathrm{unif}(0.1, 0.5)$ and $\mathrm{unif}(0.01, 0.1)$, respectively. For mixed variants (Scenarios III, VI, IX), we consider both common and rare variants with equal proportions. To simulate genotype data for unrelated individuals (Scenarios I, II, III), we first generate a latent continuous random sample for each individual from a multivariate normal distribution with mean $\mathbf{0}$ and compound symmetric covariance matrix $\boldsymbol{\Omega} = (1-\eta)\boldsymbol{I} + \eta\mathbf{1}\mathbf{1}^T$. These latent samples are then dichotomized by thresholding according to the variants' MAFs to form binary haplotypes. Finally the genotype data $\boldsymbol{G}$ are obtained by adding two independent haplotypes, inducing an LD structure that depends on the prespecified parameter η. Note, the LD covariance matrix indeed depends on both η and the variants' MAFs, as described in Section 2 of the supplemental article [Wu et al. (2018)]. Also, with dichotomization, the haplotype covariance matrix is no longer $\boldsymbol{\Omega}$, however, the parameter η can still be used to roughly indicate the

TABLE 1
Data generation scenarios

Individual setting	**Variant setting**		
	Common	**Rare**	**Mixed**
Unrelated	I	II	III
Related	IV	V	VI
Mixed	VII	VIII	IX

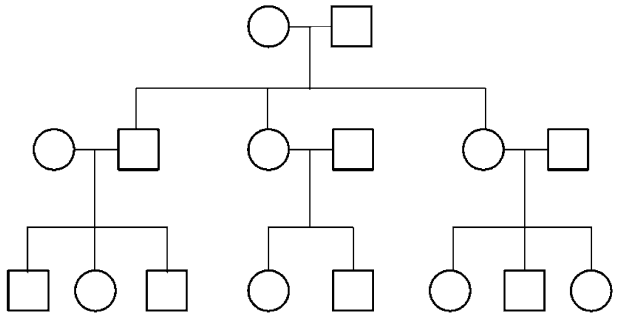

FIG. 2. *Basic family structure of* 16 *members coming from three generations, used in the simulation studies to generate data for related individuals.*

LD correlation level among variants in $\boldsymbol{G}$. We also note that this latent correlation coefficient η should not be obfuscated with the random effect parameter ρ used in Jiang and McPeek (2013), as they have obviously distinct meanings—η describes the latent correlation of the LD structure among variants in retrospective model, whereas ρ captures the heterogeneity of the random effects in prospective model.

Simulations for related individuals (Scenarios IV, V, VI) are based on a pedigree configuration with 100 outbred, 3-generation families, each containing 16 individuals related as in Figure 2. In order to simulate $\boldsymbol{G}$ with correlations across both samples and variants, as described above, we first generate multiple variant genotype data independently for founders. The genotype data for nonfounders are then generated by Mendelian "gene-dropping" along generations, assuming no recombination within haplotypes. In mixed individual setting (Scenarios VII, VIII, IX), the samples are drawn from 80 families (related as in Figure 2) and 320 unrelated individuals.

The quantitative trait data are generated from model (2.9). We assume that the design matrix $\boldsymbol{Z}$ includes the intercept and a covariate sampled independently from the standard normal distribution, with the corresponding coefficient vector $\boldsymbol{\beta} = (1, 0.6)^T$. For the variants in the genotype data $\boldsymbol{G}$, we determine their genetic effects (denoted by $\boldsymbol{\gamma}$, a vector of length m) in each scenario. In the simulation of type I error, we set $\boldsymbol{\gamma} = \boldsymbol{0}$. For the simulation of power, the genetic effects are generated under four different compositions of risk/protective/neutral variants, where the proportion of risk to protective varies from balanced (1 : 1) to unbalanced (2 : 1). The covariance of the quantitative trait $\boldsymbol{Y}$ (or the error term $\boldsymbol{\varepsilon}$) takes a variance-components form $\sigma_e^2 \boldsymbol{I} + \sigma_a^2 \boldsymbol{\Phi}$, where σ_e^2 represents variance due to random measurement error and σ_a^2 stands for variance attributed to additive polygenic random effects. The settings of these parameters σ_e^2, σ_a^2, and $\boldsymbol{\gamma}$ in the simulations are listed in Table 2. For better illustration, we set larger magnitudes for $\boldsymbol{\gamma}$ in the rare variant Scenarios (II, V, VIII) to ensure enough power.

TABLE 2
Parameter settings of variance components and genetic effects in simulations

Scenario	σ_e^2	σ_a^2	R/P/N under H_a	Settings of $\boldsymbol{\gamma}$ under H_a	
I, III	10	0	45/45/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.2)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.2, -0.05)$
			50/40/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.175)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.23, -0.05)$
			55/35/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.155)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.271, -0.05)$
			60/30/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.138)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.325, -0.05)$
II	10	0	45/45/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.5)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.5, -0.05)$
			50/40/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.445)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.569, -0.05)$
			55/35/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.4)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.657, -0.05)$
			60/30/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.363)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.775, -0.05)$
IV, VI,	2	2	45/45/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.2)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.2, -0.05)$
VII, IX			50/40/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.175)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.23, -0.05)$
			55/35/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.155)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.271, -0.05)$
			60/30/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.138)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.325, -0.05)$
V, VIII	2	2	45/45/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.5)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.5, -0.05)$
			50/40/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.445)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.569, -0.05)$
			55/35/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.4)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.657, -0.05)$
			60/30/10	$\boldsymbol{\gamma}_R \overset{\text{i.i.d.}}{\sim} \text{unif}(0.05, 0.363)$,	$\boldsymbol{\gamma}_P \overset{\text{i.i.d.}}{\sim} \text{unif}(-0.775, -0.05)$

Note: Variance components σ_e^2 and σ_a^2 represent variances attributed to random measurement error and additive polygenic random effects in the phenotypic model. Under H_a, each scenario contains four settings of genetic variants depending on the number of risk, protective, and neutral variants (R/P/N). The genetic effects of risk and protective variants are denoted by $\boldsymbol{\gamma}_R$ and $\boldsymbol{\gamma}_P$ respectively.

3.2. *Assessment of type I error.* To assess type I error, we generate 10,000 simulated data replicates from the trait model (2.9) under the null $\boldsymbol{\gamma} = \mathbf{0}$, for each scenario and each combination of $m = 10, 50, 100$ and $\eta = 0.2, 0.5, 0.8$. From the association testing results by ABT, we obtain the empirical type I error rates at nominal levels 0.01 and 0.05, as presented in Table 3. Under all scenarios and for all nine combinations of m and η, we observe that the empirical type I error rates of ABT are not significantly different from the nominal, based on a z-test at level 0.05. This shows that ABT is able to correctly control the type I error.

3.3. *Conservativeness of ABT and analysis of eigenvalues.* It is worth noting that the type I error rates presented in Table 3 are calculated from the practical null distribution of ABT instead of from the theoretical null distribution. In practice, we found that when genotypic correlations across both samples and variants exist,

TABLE 3
Empirical type I error of ABT under various scenarios and nominal levels

m	η	I	II	III	IV	V	VI	VII	VIII	IX
					Nominal level = 0.01					
10	0.2	0.0105	0.0106	0.0093	0.0109	0.0100	0.0108	0.0087	0.0090	0.0099
	0.5	0.0112	0.0094	0.0102	0.0098	0.0115	0.0092	0.0108	0.0091	0.0087
	0.8	0.0111	0.0094	0.0096	0.0100	0.0083	0.0104	0.0087	0.0089	0.0083
50	0.2	0.0103	0.0089	0.0100	0.0090	0.0097	0.0109	0.0089	0.0091	0.0087
	0.5	0.0085	0.0104	0.0099	0.0106	0.0088	0.0083	0.0094	0.0080	0.0098
	0.8	0.0102	0.0092	0.0094	0.0089	0.0095	0.0080	0.0083	0.0089	0.0086
100	0.2	0.0092	0.0082	0.0083	0.0089	0.0087	0.0081	0.0097	0.0080	0.0089
	0.5	0.0088	0.0087	0.0082	0.0102	0.0085	0.0082	0.0086	0.0092	0.0098
	0.8	0.0087	0.0088	0.0090	0.0087	0.0096	0.0087	0.0085	0.0080	0.0085
					Nominal level = 0.05					
10	0.2	0.0490	0.0512	0.0499	0.0499	0.0517	0.0509	0.0518	0.0473	0.0525
	0.5	0.0475	0.0495	0.0492	0.0487	0.0497	0.0497	0.0505	0.0504	0.0513
	0.8	0.0500	0.0484	0.0488	0.0521	0.0467	0.0513	0.0499	0.0474	0.0487
50	0.2	0.0510	0.0484	0.0497	0.0461	0.0476	0.0475	0.0468	0.0462	0.0473
	0.5	0.0483	0.0500	0.0465	0.0459	0.0478	0.0504	0.0472	0.0465	0.0484
	0.8	0.0487	0.0493	0.0486	0.0487	0.0482	0.0459	0.0480	0.0466	0.0468
100	0.2	0.0480	0.0487	0.0473	0.0474	0.0468	0.0475	0.0477	0.0466	0.0476
	0.5	0.0460	0.0460	0.0458	0.0501	0.0462	0.0462	0.0462	0.0482	0.0481
	0.8	0.0458	0.0479	0.0479	0.0481	0.0496	0.0465	0.0469	0.0490	0.0469

Note: The empirical type I error rate is calculated based on 10,000 simulated genotype replicates under the null hypothesis. The large sample 95% CIs for nominals 0.01 and 0.05 are [0.0080, 0.0120] and [0.0457, 0.0543], respectively.

using the theoretical null distribution of χ^2_m tends to make ABT conservative. We include the type I error results for ABT based on χ^2_m in Section S.3 of the supplemental article [Wu et al. (2018)]. There, we observe smaller empirical type I error rates with increasing η or by including related individuals in the sample (Scenarios IV–IX). A sensible explanation for this phenomenon may be attributed to the estimation bias of $\boldsymbol{DRD}$. Comparing equations (2.8) and (2.2) reveals that ABT is essentially a generalized, multiple-variant MASTOR test; when $\boldsymbol{R}$ is known, the m individual score statistics from the decorrelated genotype matrix $\boldsymbol{G}(\widehat{\boldsymbol{D}}\boldsymbol{R}\widehat{\boldsymbol{D}})^{-1/2}$ are independent and hence yield a χ^2_m null distribution. However, in practice (and also in our simulations), when the complex genotypic covariance structure does not follow a Kronecker product form and the LD covariance needs to be estimated, the across-column sample covariance of $\boldsymbol{\Phi}^{-1/2}\boldsymbol{G}$ may not provide an accurate estimation of $\boldsymbol{DRD}$. Hence, the across-column covariance matrix of $\boldsymbol{G}(\widehat{\boldsymbol{D}}\widehat{\boldsymbol{R}}\widehat{\boldsymbol{D}})^{-1/2}$ is not an identity matrix, though it is nearly identity for some cases, for example, Scenarios I, II, and III with unrelated individuals only. The residual across-column

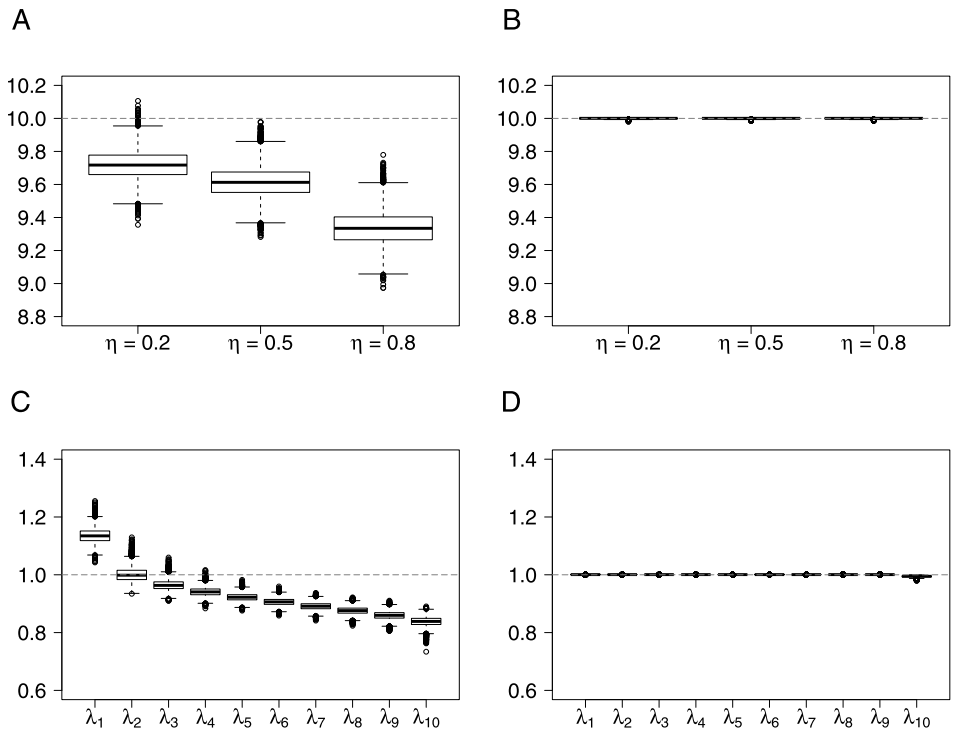

FIG. 3. *Variability of eigenvalues in Scenarios* IX *and* I. *Panel* (A): *Box plots of sum of eigenvalues from* 10,000 *simulated replicates in scenario* IX *for* $m = 10$ *and* $\eta = 0.2, 0.5, 0.8$; *Panel* (B): *Box plots of sum of eigenvalues from* 10,000 *simulated replicates in scenario* I *for* $m = 10$ *and* $\eta = 0.2, 0.5, 0.8$; *Panel* (C): *Box plots of individual eigenvalues from* 10,000 *simulated replicates in scenario* IX *for* $m = 10$ *and* $\eta = 0.8$; *Panel* (D): *Box plots of individual eigenvalues from* 10,000 *simulated replicates in scenario* I *for* $m = 10$ *and* $\eta = 0.8$.

correlation among the transformed variants results in the practical null distribution $\sum_{i=1}^{m} \lambda_i \chi^2_{1,i}$ with $\sum_{i=1}^{m} \lambda_i \leq m$, which causes the conservativeness of ABT based on χ^2_m.

To better illustrate how the complex covariance structure in genotype data influences the eigenvalues λ_i of the matrix $\boldsymbol{W}^{\#}\boldsymbol{G}^T\boldsymbol{P}\boldsymbol{G}\boldsymbol{W}^{\#}$, we present the box plots of $\sum_{i=1}^{m} \lambda_i$ in Figure 3 [see panels (A) and (B)], generated from 10,000 simulated replicates for $m = 10$ and $\eta = 0.2, 0.5, 0.8$ following Scenarios IX and I, respectively. We observe that for Scenario IX, where complex correlations exist, $\sum_{i=1}^{m} \lambda_i$ deviates from 10 and shows a decreasing trend as η increases [Figure 3, panel (A)] due to the residual correlation among the transformed variants. In contrast, for Scenario I where only LD correlation exists, $\boldsymbol{DRD}$ can be accurately estimated, and hence the individual score statistics can be treated as independent leading to the ABT null distribution approximated as χ^2_m. This can be seen from $\sum_{i=1}^{m} \lambda_i \approx m$ in Figure 3, panel (B). These results provide a reasonable explanation to the con-

servativeness of ABT based on χ^2_m, as observed from the additional type I error results in Section 3 of the supplemental article [Wu et al. (2018)]. In Figure 3, panels (C) and (D), we further present box plots for individual λ_i's (ordered from large to small) generated from the 10,000 simulated replicates for $m = 10$ and $\eta = 0.8$ from Scenarios IX and I, respectively. When complex genotypic correlations exist (Scenario IX with $\eta = 0.8$), the individual λ_i's disperse around 1 with a majority less than 1. On the other hand, when only LD correlation exists (Scenario I with $\eta = 0.8$), all individual λ_i's are very close to 1. This shows that the across-column covariance matrix of $\boldsymbol{G}(\widehat{\boldsymbol{D}}\widehat{\boldsymbol{R}}\widehat{\boldsymbol{D}})^{-1/2}$ is nearly an identity matrix in this case.

3.4. *Power comparison.* To assess power, we simulate 1000 data replicates using parameter settings described in Table 2, for each scenario with $m = 100$ and $\eta = 0.2, 0.5, 0.8$. Here, we use equal weights for the FBT, KT, and MONSTER statistics. The MONSTER test is constructed on a grid of 11 equally spaced points—$\rho_1 = 0, \rho_2 = 0.1, \ldots, \rho_{10} = 0.9, \rho_{11} = 1$, as in its original paper [Jiang and McPeek (2013)]. The empirical power of FBT, KT, MONSTER, and ABT for typical scenarios (II, IV, V, IX) at $\alpha = 0.05$ are shown in Figure 4 (the case of $\alpha = 0.01$ is included in Section S.4 of the supplemental article [Wu et al. (2018)]). Compared to the other three tests, ABT achieves considerably higher power under different settings of η and $\boldsymbol{\gamma}$ in all these scenarios. In particular, we observe that, due to the coexistence of risk and protective variants (though their relative proportion varies), FBT loses power, and KT and MONSTER have similar and overall good performance. Under weak LD settings ($\eta = 0.2$), ABT has comparable power to KT and MONSTER, whereas when η is moderate (0.5) or large (0.8), ABT has higher power than others. This is as expected since ABT is able to incorporate the LD covariance information using retrospective model whereas other methods cannot.

3.5. *Adaptability of weights $\boldsymbol{W}^*$ to the direction of genetic effects.* One obvious advantage of ABT over other fixed-weight tests is that the weights $\boldsymbol{W}^*$ adapt to the direction of true genetic effects. In Section 2.3, we have shown that $\boldsymbol{W}^*$ is able to maximize the test statistic of FBT. In order to understand how $\boldsymbol{W}^*$ can help gain power in contrast to prescribed weights, we compare the signs of $\boldsymbol{W}^*$ to those of the genetic effects $\boldsymbol{\gamma}$ using the simulated data sets in the power analysis. Figure 5, panel (A), presents box plots of the weights $\boldsymbol{W}^*$ across the 1000 replicates in Scenario IV for $m = 40$, $\eta = 0.8$ with balanced setting of $\boldsymbol{\gamma}$. We note that, under this setting, the first 45% components of $\boldsymbol{\gamma}$ are positive followed by the next 45% being negative, and the remaining 10% are zeros. This box plot clearly demonstrates that on average, the weights $\boldsymbol{W}^*$ are able to track the direction of true genetic effects resulting in stronger association of the weighted sum genetic score. On the contrary, if one adopts FBT with the Madsen–Browning weights (which are all positive), the effects from risk and protective variants will be canceled through linear combination, thereby weakening the association of the weighted sum genetic

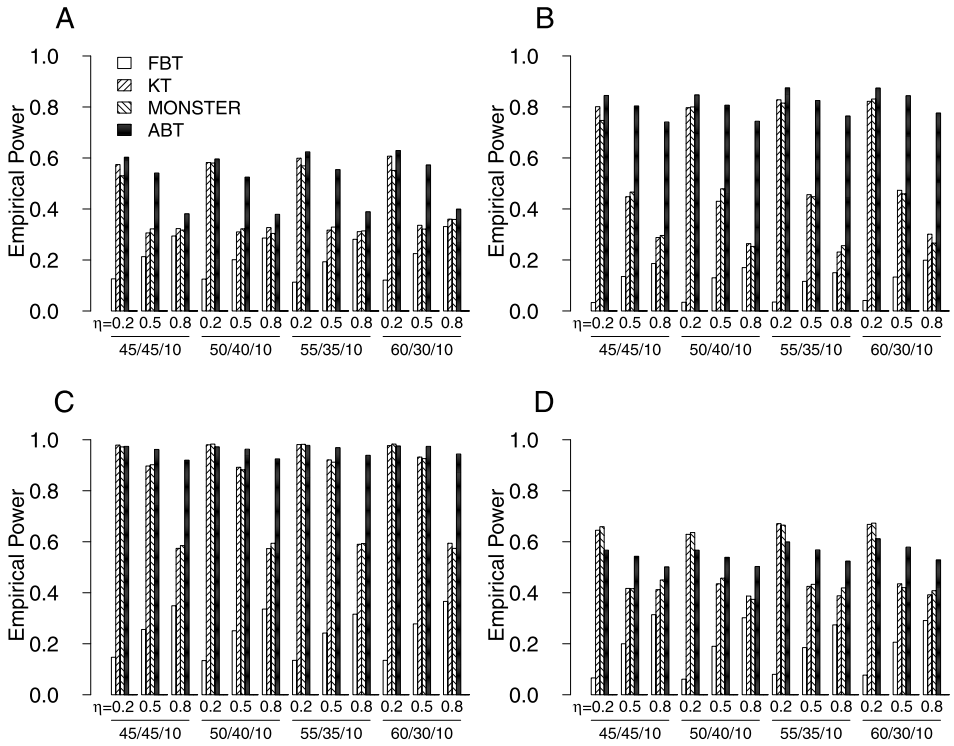

FIG. 4. *Empirical power of FBT, KT, MONSTER, and ABT at* $\alpha = 0.05$, *based on* 1000 *simulated replicates with* $m = 100$. *Panel* (A): *Scenario* II; *Panel* (B): *Scenario* IV; *Panel* (C): *Scenario* V; *Panel* (D): *Scenario* IX.

score. Figure 5, panel (B), provides the average of mismatch rates (i.e., proportion of $\boldsymbol{\gamma} \times \boldsymbol{W}^* < 0$ over the causal variants, where $\times$ denotes component-wise product of two vectors) across the 1000 replicates in Scenarios II, IV, V, and IX for $m = 100$, η varying from 0 to 0.9, and under balanced setting of $\boldsymbol{\gamma}$. We observe that for these scenarios, the average mismatch rates are about 22%–34% and slowly increasing with η. Intuitively, lower mismatch rates are indicative of better adaptability of $\boldsymbol{W}^*$ leading to higher power. This intuition is verified from the constant lower average mismatch rate in Scenarios IV and V [lines with circle and triangle markers compared to lines with square and diamond markers in Figure 5, panel (B)] and the consistent higher power of ABT in Scenarios IV and V [panels (B) and (C) in Figures 4 compared to panels (A) and (D)].

4. Application: Association analysis of the GO-ESP data. The NHLBI "Grand Opportunity" Exome Sequencing Project (GO-ESP) is a study for identifying genetic variants in coding regions (exons) of the human genome that are associated with heart, lung, and blood diseases. By pioneering the application of next

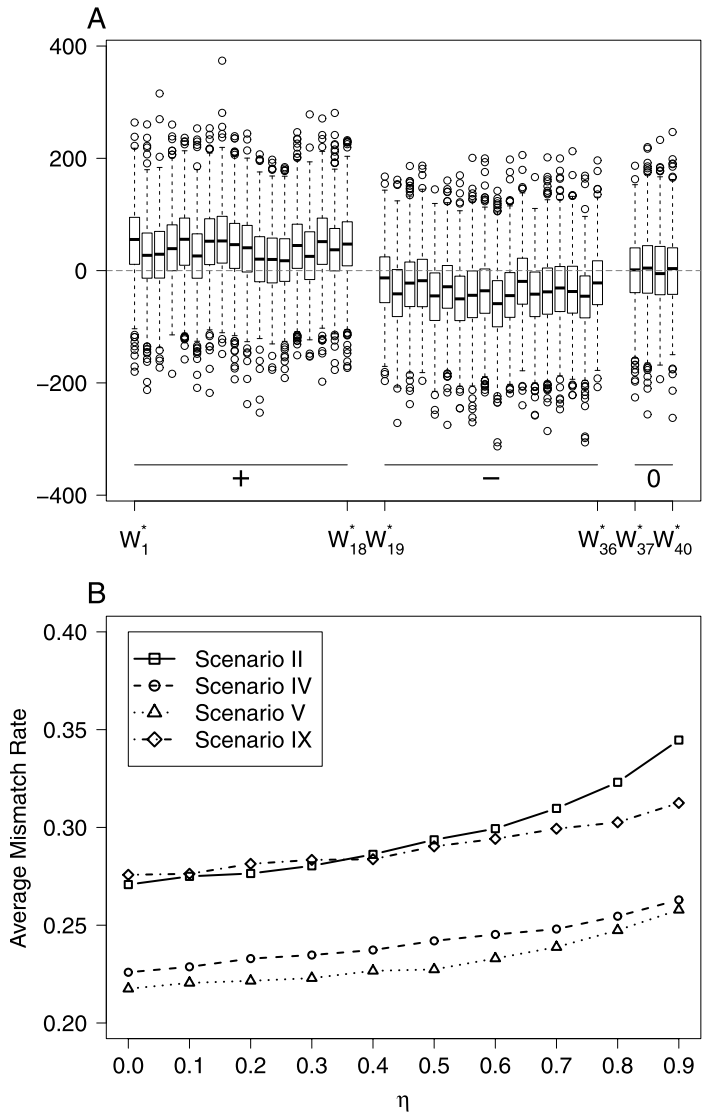

FIG. 5. *Illustration of adaptability of* $\boldsymbol{W}^*$ *to the direction of true genetic effects. Panel* (A): *Box plots of weights* $\boldsymbol{W}^*$ *across* 1000 *simulated replicates in Scenario* IV *for* $m = 40$ *and* $\eta = 0.8$ *with balanced setting of* $\boldsymbol{\gamma}$; *Panel* (B): *Average mismatch rates across* 1000 *simulated replicates in Scenarios* II, IV, V, *and* IX *for* $m = 100$, η *varying from* 0 *to* 0.9, *and under balanced setting of* $\boldsymbol{\gamma}$.

generation exome sequencing across diverse, richly phenotyped populations, this project aims to discover novel genes and mechanisms contributing to heart, lung, and blood disorders. Our use of the GO-ESP data was approved by the Institutional Review Board of Virginia Tech. In the GO-ESP, a total of 499 Framingham Heart

Study (FHS) participants were selected for exome sequencing, leading to 460 finished sequence data in dbGaP (dbGaP Study Accession: phs000401.v12.p10). Individual level phenotype data were collected for 463 participants, including the fasting glucose measurements at multiple time points (some individuals are from cohort 2 of the FHS with at most eight measurements, whereas others are from cohort 3 with at most two measurements). Our association analysis treats the fasting glucose level at first visit as the quantitative trait. To avoid noise and have results comparable to other fasting glucose studies, type 2 diabetics (indicated by ESP_t2diabetes_baseline = 1) and individuals with fasting glucose > 125 mg/dl were excluded from our analysis. Among the remaining 427 phenotyped individuals, 173 are from 64 families, and the rest are unrelated individuals. We then adjusted the quantitative trait for age and sex and tested its association with the gene regions located on all 22 chromosomes.

Table 4 reports ten top-ranked p-values for testing the association between fasting glucose level and gene regions on all 22 chromosomes based on the GO-ESP data by using ABT. For comparison, the FBT, KT, and MONSTER p-values are also listed. Among these genes, *CYP1A2* encodes a member of the cytochrome P450 superfamily of enzymes which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids, and other lipids. Variants from this gene were found to affect the risk of impaired fasting glucose in hypertension caused by coffee consumption [Palatini et al. (2015)], show associations with higher serum clozapine concentrations and an increased risk of developing

TABLE 4
P-values for association of fasting glucose level with gene regions on 22 chromosomes based on the GO-ESP data

Gene	Chr	# genotyped SNPs	p-value based on ABT	FBT	KT	MONSTER
MIR411	14	2	3.51×10^{-5}	5.91×10^{-2}	3.19×10^{-2}	3.28×10^{-2}
LINC00472	6	5	8.55×10^{-5}	4.60×10^{-1}	8.18×10^{-2}	8.74×10^{-2}
KRTDAP	19	10	2.19×10^{-4}	4.02×10^{-1}	1.19×10^{-4}	2.10×10^{-4}
FXYD6-FXYD2	11	19	3.70×10^{-4}	6.56×10^{-1}	1.43×10^{-3}	1.12×10^{-3}
GHITM	10	9	4.11×10^{-4}	1.26×10^{-1}	1.28×10^{-5}	3.38×10^{-5}
CYP1A2	15	15	4.78×10^{-4}	1.51×10^{-2}	3.76×10^{-5}	2.69×10^{-4}
GIPR	19	12	5.69×10^{-4}	4.27×10^{-1}	1.16×10^{-1}	1.25×10^{-1}
SRRM3	7	15	6.08×10^{-4}	1.27×10^{-1}	1.51×10^{-1}	1.54×10^{-1}
CDKL3	5	12	6.18×10^{-4}	2.22×10^{-1}	5.90×10^{-2}	5.89×10^{-2}
FST	5	8	7.39×10^{-4}	8.56×10^{-1}	1.38×10^{-2}	1.15×10^{-2}

Note: The quantitative trait—fasting glucose level was adjusted for age and sex. For FBT, KT, and MONSTER, Madsen–Browning weights were used. MIM numbers of genes: *KRTDAP*[617212], *FXYD6*[606683], *FXYD2*[601814], *CYP1A2*[124060], *GIPR*[137241], *CDKL3*[608459], *FST*[136470].

insulin and lipid elevations and insulin resistance on a given dose of clozapine [Melkersson et al. (2007)]. The glucose-dependent insulinotropic polypeptide receptor (GIPR) gene has been extensively studied and demonstrated to stimulate insulin release in the presence of elevated glucose. *GIPR* was found to play a role in the incretin effect and in early pathophysiologic pathways that could lead to impaired glucose tolerance and type 2 diabetes in humans [Saxena et al. (2010)]. More findings about this gene regarding regulating glucose and insulin metabolism in humans were reported in recent GWASs [Ingelsson et al. (2010), Qi et al. (2012)]. The *FST* gene is an important regulator of activin, which might have important roles in insulin resistance and the onset and development of type 2 diabetes [Hansen et al. (2013), Wu et al. (2012)].

5. Conclusions. The burden tests and the kernel tests are supposedly the two major classes of methods for multiple-variant association analyses that are widely used in genetic association studies. Though each has its advantages, there are two deficiencies shared in common in their implementation to genotype data with complex correlations:

(i) Existing methods in both classes are mostly developed under the prospective regression setting (focusing on characterizing the conditional expectation of random trait measurements given covariates and genotypes), where accounting for LD correlations among variants is not straightforward.

(ii) The weights adopted in both classes are usually prescribed, which are not adaptable to the direction of individual variant effects, and therefore may not achieve optimal performance for association testing.

In view of these issues, we first develop a retrospective, fixed-weight burden test to incorporate genotypic correlations across both samples and variants, and then employ data-driven weights to maximize the statistic of this fixed-weight burden test. The resulting Adaptive-weight Burden Test can be easily constructed by first projecting genotype data to eliminate LD correlations and then combining independent MASTOR tests on the transformed variants. The ABT testing method sheds light on a number of aspects as described below.

First, by using a retrospective setting and treating genotype data as random, the ABT is able to directly model correlations across both samples and variants which exist universally in genotype data collected for current genetic association studies. This is highly desirable because very few existing methods can make full use of the valuable genotypic correlation information arising from both Mendelian inheritance and nonrandom association of alleles at different loci. A literature search revealed that the MONSTER test is perhaps the closest method accounting for the complex genetic correlations. However, because of its prospective setting, MONSTER cannot model LD correlation among variants directly, though this information can be thought of being implicitly carried through formulating the covariance $\boldsymbol{R}_\rho$ of the variant random effects vector $\boldsymbol{\beta}$. In addition, its application may

be restricted by the compound symmetric assumption of the covariance structure $\boldsymbol{R}_\rho$. Moreover, modeling phenotypes as fixed is also theoretically appealing because it makes fewer assumptions about the phenotypic covariance structure [Price et al. (2010b)]. These facts clearly show the advantage of retrospective modeling to prospective modeling in genetic association analysis.

Second, ABT adopts data-driven weights in the collapsing procedure. Unlike many other existing weighting strategies for burden tests or kernel tests, which lack theoretical justification, these weights guarantee ABT to be statistically powerful. The calculation of the data-driven weights is straightforward, and the resulting ABT statistic is shown to have an explicit null distribution. Through extensive simulations, we demonstrated that ABT is able to control type I error and achieves higher power than the other three competing methods FBT, KT, and MONSTER in almost all scenarios with moderate or large LD correlation. Further investigation reveals that the weights of ABT are able to adapt to the direction of true genetic effects. This overcomes the main drawback of fixed-weight burden tests, which lose power in the presence of both risk and protective variants.

Third, although ABT is derived from a burden test perspective, we showed that it is formally equivalent to a family-based kernel test with the GMB weights. This interesting finding can be used to guide the selection of weights in traditional kernel tests to accommodate LD correlation among variants. It also motivates statistical geneticists to reconsider and to explore in-depth the relation between burden tests and kernel tests. Our simulations demonstrate that, when genotypic correlations exist across both samples and variants, using the theoretical null distribution of χ^2_m results in a conservative test. A plausible explanation for this may be attributed to the covariance estimation bias in the decorrelation procedure. Additional eigenvalue analysis in our simulation study confirms this conjecture. Hence, we suggest using the practical null distribution (mixture of χ^2_1's) when complex genotypic correlations are not separable and when LD correlation $\boldsymbol{R}$ needs to be estimated from genotype data $\boldsymbol{G}$.

Finally, as a retrospective association test, ABT is expected to have several additional advantages, for example, in borrowing information from partially informative data and in maintaining robustness to phenotype model misspecification. The present work is an initial study toward a complete exploration on the underpinning and characteristics of this newly developed method. As more and more high-throughput sequencing data on samples with complex correlation structure become available in recent GWASs, it is of significant importance to investigate the performance of ABT in detecting association for rare variants and for samples with population or cryptic relatedness structure. This will be pursued elsewhere.

Acknowledgments. The authors would like to thank the Editor, the Associate Editor, and the Reviewers, whose constructive and insightful comments contributed to a significantly improved version of this article.

SUPPLEMENTARY MATERIAL

Mathematical justifications and additional results (DOI: 10.1214/17-AOAS1121SUPP; .pdf). The supplementary materials of the paper are organized as follows. Supplement S.1 provides the theoretical justification of the covariance matrix of genetic burden score $\boldsymbol{X}$. Supplement S.2 derives the LD covariance of the simulated genotype data for founders. In Supplement S.3, additional results from Section 3.3 on the empirical type-I error of ABT based on χ^2_m null distribution in simulation studies are summarized in Table S1. Supplement S.4 includes additional power comparison results at $\alpha = 0.01$, for Scenarios II, IV, V, and IX.

REFERENCES

ANSORGE, W. J. (2009). Next-generation DNA sequencing techniques. *New Biotechnol.* **25** 195–203.

ASIMIT, J. and ZEGGINI, E. (2010). Rare variant association analysis methods for complex traits. *Annu. Rev. Genet.* **44** 293–308.

CHEN, H., MEIGS, J. B. and DUPUIS, J. (2013). Sequence kernel association test for quantitative traits in family samples. *Genet. Epidemiol.* **37** 196–204.

CHUN, H., BALLARD, D. H., CHO, J. and ZHAO, H. (2011). Identification of association between disease and multiple markers via sparse partial least-squares regression. *Genet. Epidemiol.* **35** 479–486.

FANG, S., ZHANG, S. and SHA, Q. (2014). Detecting association of rare variants by testing an optimally weighted combination of variants for quantitative traits in general families. *Ann. Hum. Genet.* **77** 524–534.

FUENTES, M. (2006). Testing for separability of spatial-temporal covariance functions. *J. Statist. Plann. Inference* **136** 447–466.

GAUDERMAN, W. J., MURCRAY, C., GILLILAND, F. and CONTI, D. V. (2007). Testing association between disease and multiple SNPs in a candidate gene. *Genet. Epidemiol.* **31** 383–395.

HAN, F. and PAN, W. (2010). A data-adaptive sum test for disease association with multiple common or rare variants. *Hum. Hered.* **70** 42–54.

HANSEN, J., RINNOV, A., KROGH-MADSEN, R., FISCHER, C. P., ANDREASEN, A. S., BERG, R. M., MØLLER, K., PEDERSEN, B. K. and PLOMGAARD, P. (2013). Plasma follistatin is elevated in patients with type 2 diabetes: Relationship to hyperglycemia, hyperinsulinemia, and systemic low-grade inflammation. *Diabetes/Metab. Res. Rev.* **29** 463–472.

INGELSSON, E., LANGENBERG, C., HIVERT, M. F., PROKOPENKO, I., LYSSENKO, V., DUPUIS, J., MÄGI, R., SHARP, S., JACKSON, A. U., ASSIMES, T. L. et al. (2010). Detailed physiologic characterization reveals diverse mechanisms for novel genetic loci regulating glucose and insulin metabolism in humans. *Diabetes* **59** 1266–1275.

JAKOBSDOTTIR, J. and MCPEEK, M. S. (2013). MASTOR: Mixed-model association mapping of quantitative traits in samples with related individuals. *Am. J. Hum. Genet.* **92** 652–666.

JIANG, D. and MCPEEK, M. S. (2013). Robust rare variant association testing for quantitative traits in samples with related individuals. *Genet. Epidemiol.* **38** 1–20.

KWEE, L. C., LIU, D., LIN, X., GHOSH, D. and EPSTEIN, M. P. (2008). A powerful and flexible multilocus association test for quantitative traits. *Am. J. Hum. Genet.* **82** 386–397.

LEE, S., WU, M. C. and LIN, X. (2012). Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* **13** 762–775.

LEE, S., EMOND, M. J., BAMSHAD, M. J., BARNES, K. C., RIEDER, M. J., NICKERSON, D. A., NHLBI GO EXOME SEQUENCING PROJECT-ESP LUNG PROJECT TEAM, CHRISTIANI, D. C., WURFEL, M. M. and LIN, X. (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.* **91** 224–237.

LI, B. and LEAL, S. M. (2008). Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data. *Am. J. Hum. Genet.* **83** 311–321.

LI, M. X., GUI, H. S., KWAN, J. S. and SHAM, P. C. (2011). GATES: A rapid and powerful gene-based association test using extended simes procedure. *Am. J. Hum. Genet.* **88** 283–293.

LIN, D. Y. and TANG, Z. Z. (2011). A general framework for detecting disease associations with rare variants in sequencing studies. *Am. J. Hum. Genet.* **89** 354–367.

LIU, D. J. and LEAL, S. M. (2010). A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. *PLoS Genet.* **6** e1001156.

MA, L., CLARK, A. G. and KEINAN, A. (2013). Gene-based testing of interactions in association studies of quantitative traits. *PLoS Genet.* **9** e1003321.

MADSEN, B. E. and BROWNING, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* **5** e1000384.

MCPEEK, M. S., WU, X. and OBER, C. (2004). Best linear unbiased allele-frequency estimation in complex pedigrees. *Biometrics* **60** 359–367.

MELKERSSON, K. I., SCORDO, M. G., GUNES, A. and DAHL, M.-L. (2007). Impact of CYP1A2 and CYP2D6 polymorphisms on drug metabolism and on insulin and lipid elevations and insulin resistance in clozapine-treated patients. *J. Clin. Psychiatry* **68** 697–704.

MORGENTHALER, S. and THILLY, W. G. (2007). A strategy to discover genes that carrymulti-allelic or mono-allelic risk for common diseases: A Cohort Allelic Sums Test (CAST). *Mutation Research* **615** 28–56.

NEALE, B. M., RIVAS, M. A., VOIGHT, B. F., ALTSHULER, D., DEVLIN, B., ORHO-MELANDER, M., KATHIRESAN, S., PURCELL, S. M., ROEDER, K. and DALY, M. J. (2011). Testing for an unusual distribution of rare variants. *PLoS Genet.* **7** e1001322.

PALATINI, P., BENETTI, E., MOS, L., GARAVELLI, G., MAZZER, A., COZZIO, S., FANIA, C. and CASIGLIA, E. (2015). Association of coffee consumption and CYP1A2 polymorphism with risk of impaired fasting glucose in hypertensive patients. *Eur. J. Epidemiol.* **30** 209–217.

PRICE, A. L., KRYUKOV, G. V., DE BAKKER, P. I., PURCELL, S. M., STAPLES, J., WEI, L. J. and SUNYAEV, S. R. (2010a). Pooled association tests for rare variants in exon-resequencing studies. *Am. J. Hum. Genet.* **86** 832–838.

PRICE, A. L., ZAITLEN, N. A., REICH, D. and PATTERSON, N. (2010b). New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11** 459–463.

QI, Q., BRAY, G. A., HU, F. B., SACKS, F. M. and QI, L. (2012). Weight-loss diets modify glucose-dependent insulinotropic polypeptide receptor rs2287019 genotype effects on changes in body weight, fasting glucose, and insulin resistance: The preventing overweight using novel dietary strategies trial. *Am. J. Clin. Nutr.* **95** 506–513.

SAXENA, R., HIVERT, M. F., LANGENBERG, C., TANAKA, T., PANKOW, J. S., VOLLENWEIDER, P., LYSSENKO, V., BOUATIA-NAJI, N., DUPUIS, J., JACKSON, A. U. et al. (2010). Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* **42** 142–148.

SCHAID, D. J., MCDONNELL, S. K., SINNWELL, J. P. and THIBODEAU, S. M. (2013). Multiple genetic variant association testing by collapsing and kernel methods with pedigree or population structured data. *Genet. Epidemiol.* **37** 409–418.

SCHIFANO, E. D., EPSTEIN, M. P., BIELAK, L. F., JHUN, M. A., KARDIA, S. L. R., PEYSER, P. A. and LIN, X. (2012). SNP set association analysis for familial data. *Genet. Epidemiol.* **36** 797–810.

SHA, Q. and ZHANG, S. (2014). A novel test for testing the optimally weighted combination of rare and common variants based on data of parents and affected children. *Genet. Epidemiol.* **38** 135–143.

SHA, Q., WANG, X., WANG, X. and ZHANG, S. (2012). Detecting association of rare and common variants by testing an optimally weighted combination of variants. *Genet. Epidemiol.* **36** 561–571.

SHENDURE, J. and JI, H. (2008). Next-generation DNA sequencing. *Nat. Biotechnol.* **26** 1135–1145.

SPLANSKY, G. L., COREY, D., YANG, Q., ATWOOD, L. D., CUPPLES, L. A., BENJAMIN, E. J., D'AGOSTINO SR., R. B., FOX, C. S., LARSON, M. G., MURABITO, J. M. et al. (2007). The third generation cohort of the national heart, lung, and blood institute's framingham heart study: Design, recruitment, and initial examination. *Am. J. Epidemiol.* **165** 1328–1335.

THE 1000 GENOMES PROJECT CONSORTIUM (2010). A map of human genome variation from population-scale sequencing. *Nature* **467** 1061–1073.

THORNTON, T. and MCPEEK, M. S. (2007). Case-control association testing with related individuals: A more powerful quasi-likelihood score test. *Am. J. Hum. Genet.* **81** 321–337.

THORNTON, T. and MCPEEK, M. S. (2010). ROADTRIPS: Case-control association testing with partially or completely unknown population and pedigree structure. *Am. J. Hum. Genet.* **86** 172–184.

WANG, K. and ABBOTT, D. (2008). A principal components regression approach to multilocus genetic association studies. *Genet. Epidemiol.* **32** 108–118.

WANG, Y., CHEN, Y. H. and YANG, Q. (2012). Joint rare variant association test of the average and individual effects for sequencing studies. *PLoS ONE* **7** e32485.

WANG, T. and ELSTON, R. C. (2007). Improved power by use of a weighted score test for linkage disequilibrium mapping. *Am. J. Hum. Genet.* **80** 353–360.

WANG, X., MORRIS, N. J., ZHU, X. and ELSTON, R. C. (2013a). A variance component based multi-marker association test using family and unrelated data. *BMC Genet.* **14** 17.

WANG, X., LEE, S., ZHU, X., REDLINE, S. and LIN, X. (2013b). GEE-based SNP set association test for continuous and discrete traits in family based association studies. *Genet. Epidemiol.* **37** 778–786.

WU, M. C., KRAFT, P., EPSTEIN, M. P., TAYLOR, D. M., CHANOCK, S. J., HUNTER, D. J. and LIN, X. (2010). Powerful SNP-set analysis for case-control genome-wide association studies. *Am. J. Hum. Genet.* **86** 929–942.

WU, M. C., LEE, S., CAI, T., LI, Y., BOEHNKE, M. and LIN, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89** 82–93.

WU, H., WU, M., CHEN, Y., ALLAN, C. A., PHILLIPS, D. J. and HEDGER, M. P. (2012). Correlation between blood activin levels and clinical parameters of type 2 diabetes. *Exp. Diabetes Res.* **2012** 410579.

WU, X., GUAN, T., LIU, D. J., NOVELO, L. G. and BANDYOPADHYAY, D. (2018). Supplement to "Adaptive-weight burden test for associations between quantitative traits and genotype data with complex correlations." DOI:10.1214/17-AOAS1121SUPP.

X. WU
T. GUAN
DEPARTMENT OF STATISTICS
VIRGINIA TECH
250 DRILLFIELD DRIVE, MC0439
BLACKSBURG, VIRGINIA 24061
USA
E-MAIL: xwwu@vt.edu
ting@vt.edu

D. J. LIU
DEPARTMENT OF PUBLIC HEALTH SCIENCES
HERSHEY INSTITUTE OF PERSONALIZED MEDICINE
PENNSYLVANIA STATE UNIVERSITY
COLLEGE OF MEDICINE
HERSHEY, PENNSYLVANIA 17033
USA
E-MAIL: dajiang.liu@psu.edu

L. G. L. NOVELO
DEPARTMENT OF BIOSTATISTICS
SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF TEXAS HEALTH
SCIENCE CENTER
HOUSTON, TEXAS 77030
USA
E-MAIL: Luis.G.LeonNovelo@uth.tmc.edu

D. BANDYOPADHYAY
DEPARTMENT OF BIOSTATISTICS
SCHOOL OF MEDICINE
VIRGINIA COMMONWEALTH UNIVERSITY
RICHMOND, VIRGINIA 23298
USA
E-MAIL: bandyop@vcuhealth.org